TRIP REPORT

GEOFFREY R. HOWE, PH.D.

VISIT TO KIEV, UKRAINE

Thursday, May 28, 1998
to
Friday, June 5, 1998

I. Introduction:

The first part of the visit, i.e., Thursday and Friday, May 28 and 29, was devoted to the leukemia project. In particular, extensive discussions were held with Dr. Natalia Gudzenko and her colleagues to discuss epidemiologic aspects of the study. In addition, discussions were held with Dr.Vadim Chumak and his colleagues, again, in consultation with Dr. Gudzenko to discuss dosimetric aspects, in particular, those which overlap with epidemiologic activities. A visit was also paid to the Chernobyl State Registry (Dr. Cortushin) to discuss some issues relating to the Registry.

The second week, i.e., Tuesday, June 2 to Friday, June 5, 1998, was devoted to the thyroid project.

II. Leukemia Project:

During the first two days of the visit, various topics were discussed at various times with different individuals. Therefore, this report is organized by topic rather than necessarily by chronology. The main topics discussed, and relevant comments and conclusions, follow.

A. Status of the Pilot Project in Dnepropetrovsk Oblast:

Both Dr. Gudzenko and Professor Burch had participated during the previous week in the field trip to Dnepropetrovsk to assess the current status of the pilot project. From discussions with them, the following points emerged.

① To date, little field work appears to have been carried out in Dnepropetrovsk, although plans have been developed for such field work and it is anticipated that such field work will start in the immediate future.

② A total of approximately 18,000 liquidators currently resident in Dnepropetrovsk oblast has been identified from the oblast or state Chernobyl registries. Liquidators in the oblast are still being added to the oblast Chernobyl Registry, including some who worked between 1986 and 1990. Last year in total there were an additional 500 liquidators registered. These, presumably, are individuals moving into the oblast from elsewhere, but it is unclear whether this is the only source of new registrants. Data on new registrants are sent from the oblast Chernobyl Registry (which is computerized) to the State Registry and are being added to the latter, although, presumably, they may already be on the State Registry if they were registered in another oblast previously. Presumably, there is a mechanism for avoiding duplication of such individuals in the State Registry but, again, the process of determining the existence of such duplicates is unclear.

③ Of the 18,000 liquidators on the State Registry currently resident in Dnepropetrovsk oblast, the great majority first worked at Chernobyl between 1986 and 1990; Dr. Gudzenko estimates that only a "few hundred" would not have worked during this period.

④ During the past year, approximately 14,000 of the 18,000 liquidators have been examined in one of the raion polyclinics. Of the 4,000 "lost to follow-up," approximately 2,000 have been seen within the past two years, 600 were lost three years ago, 300, four years ago, and 600 five or more years ago. It is anticipated that many of the 2,000 seen two years ago, but not in the past year, will reappear for a subsequent medical examination according to Dr. Gudzenko. If this is true, and the final lost to follow-up for that group is, in fact, similar to other years, i.e., about 600, then the medical examination system would be able to trace approximately between 80% and 85% of the cohort, which given that medical screening started more than 10 years ago, is encouraging. However, it appears as though the attrition rate is remaining relatively constant from year to year so

in a number of years in the future the lost to follow-up rate could become quite substantial and, thus, it will be essential that other follow-up mechanisms be available (see below).

⑤ Currently, 50 liquidators lost to follow-up three or more years ago have been identified in the State Registry and it has been confirmed at the oblast level that, indeed, these individuals are lost to follow-up, i.e., the oblast polyclinic has had no contact with them during the past three years. I recommended that the sample should be expanded to include those lost for one or two years since, of course, there is no guarantee that these individuals will return subsequently for medical examinations. I also recommended that the sample should be stratified by years of loss to follow-up and that the number in the sample in each of those years would then be proportional to the total numbers in those years. If possible, the sample size should be expanded to about 100 individuals, but if this is not possible, the sample should be expanded from 50 to the number that resources will permit.

⑥ I strongly recommended that the procedures to be used for tracing the sample of lost to follow-up liquidators should be clearly documented, and the order of priority of the various approaches to be used should be specifically delineated. In addition, of course, the result of the follow-up efforts should be recorded by the success or failure of the various procedures, and by year lost to follow-up.

⑦ One point which was not clear to me was what procedure would be employed with individuals who have been lost to follow-up but were traced during the pilot study. Ideally, at least some of these individuals should be invited to the polyclinic in order to undergo administration of the questionnaire and, ideally, a further subset would have blood drawn. At the moment, the plan appears to be to invite 40 individuals who are ***not*** lost to follow-up back to the polyclinic for questionnaire administration, with a subset of these individuals having their blood drawn. However, this latter group is likely to be more compliant than those lost to follow-up so it will be important to assess the participation rate by those lost to follow-up as well as those currently under observation.

⑧ In order to address the above point, I suggested the following sampling scheme. The first 50 liquidators should be randomly chosen from amongst those lost to follow-up. Attempts should be made to trace these individuals, and of those who are traced, a random sample of ten should be invited to come to the polyclinic for questionnaire administration and blood samples. If some of these do not turn up, the invitation could be extended to another group of traced individuals until the number is made up to ten. This would enable the administration of the questionnaire to ten individuals to be tested. A sample of 30 individuals from those ***not*** lost to follow-up should also be selected from the Chernobyl Registry and, again, invited to attend the polyclinic for interviewing and blood collection. Again, if some of these fail to turn up, substitutes should be invited until a total of 30 individuals have been interviewed. Thus, in total, 40 individuals will have received the questionnaire and blood will have been taken from all of them. However, it may be regarded as unethical to take blood from 40 individuals, but only use blood from ten individuals for further processing. If so, then the individuals whose blood will be taken will have to be randomly sampled from amongst the 40 individuals who received the questionnaire. This should provide a handle on response rates from amongst both those lost to follow-up and those not lost to follow-up, though denominators will be small and acceptance rates will be unstable statistically speaking.

B. Status of Oblasts Other than Dnepropetrovsk:

① It is planned to run tabulations for each of the other five oblasts in the State Chernobyl Registry which will do a head count of liquidators by years of loss to follow-up, i.e., not returned for medical examination, in the same way that has already been done for Dnepropetrovsk oblast. The software for this, of course, should already exist, but it seems unlikely that Dr. Cortushin and his staff would be willing to do this until the hardware from NCI is installed, which may well be a period of six months from now (see below).

② Dr. Gudzenko's "impressions" of the other five oblasts in relation to the likely quality of data in the State Registry from those oblasts are as follows:

- Donetsk and Kharkiv should be as good as Dnepropetrovsk;

- Sumskaya, nothing known;
- Kiev City appears to have problems in their database, but these problems are unknown to Dr. Gudzenko;
- Kiev oblast had problems two years ago in the sense that they sent identical data to that which had been sent the previous year; however, last year the problem appeared to have been rectified and updated data sent to the State Registry.

C. Passive Follow-up and Record Linkage:

① In addition to the active follow-up methods discussed above, I also stressed with Dr. Gudzenko issues of passive follow-up using record linkage techniques. At the moment deaths amongst liquidators are (sometimes) identified by a local physician in the polyclinic, notified to the oblast and eventually included in the Chernobyl State Registry. It is not clear how complete notification of deaths is by this method. Dr. Gudzenko said that currently death records for the oblasts are available on paper but not in computerized form. It might be feasible to conduct a manual record linkage amongst those lost to follow-up. For example, in Dnepropetrovsk if this was 15% of 18,000, i.e., 2,700 this ***might*** be feasible given the necessary resources. Exercises of this size, i.e., manual linkage have been carried out successfully in Statistics Canada using microfiche records; alternatively, it might be possible to do this on a sample basis to estimate how complete is the current system for notifying deaths.

② Some potential resources for passive follow-up other than the death records mentioned above were identified. These include the passport office (Ministry of Internal Affairs), but recently there have been some changes in passport numbers and it is unknown whether or not records of such changes have been kept and if they are computerized. Another possible resource is a "benefits file" of liquidators which, apparently, is kept in computerized form at the Ministry of Chernobyl Affairs, and, presumably, would have current addresses and possibly death information for liquidators. Another possible source is a file of "lifetime events" which, apparently, is maintained for Ukrainian citizens and contains records of births, marriages and deaths in computerized form and, apparently,

is kept by some ministry (unidentified). I strongly recommended that these various sources be investigated as soon as possible, in particular, the availability of computerized data, the content of such data records and the completeness of coverage of such files. In addition, I also recommended that, should these sources appear useful, permission to obtain access to them should be initiated as soon as possible but, clearly, this will probably have to be done at a senior level, i.e., Dr. Romanyenko's office. Investigation of the resources and obtaining of permission, where appropriate, should be carried out during the pilot phase of the study, though clearly these mechanisms cannot realistically be tested during the pilot phase.

③ As discussed above, eventually it may prove necessary to link the cohort file to passport records, the benefits file records and the "life events" records. This should only be necessary for those lost to follow-up through the medical examination system. It is hoped that these linkages can be carried out using computerized techniques if the appropriate records are indeed computerized. These linkages will almost certainly have to be carried out in the offices where the various files are held, i.e., the cohort records and record linkage system will have to be transferred from the Institute of Radiation Medicine to the relevant computer systems.

④ The other critical linkage will, of course, be that to leukemia/lymphoma records, both retrospectively and prospectively. Although, originally the plan was to make use of the cancer registry data for years when this was available, according to Dr. Gudzenko, this may not be feasible. The problem appears to be that the Institute of Radiation Medicine will not release the file of the cohort to the Cancer Registry and the Cancer Registry will not release their file of cancer incidence data to the Institute. There appears to be some rivalry between these two organizations, both of which regard themselves as research institutes and there appears to be a perception on both sides that the other side could use the data they receive in some unauthorized fashion. How real this is, and whether it can be resolved through future discussions, remains unclear. However, at the moment, it seems best to allow for the possibility that the linkages to the leukemia/lymphoma data will have

to be based on the extraction of the latter records from the oblast hematological departments and dispensaries.

⑤ In terms of implementing computerized record linkage procedures in the leukemia study, I propose that it would be much more efficient to provide training in Kiev in both the theory and practical application of such record linkage. This would take the form of a small group training session to be conducted by myself and a Russian-speaking programmer from North America who is familiar with the operation of a standard record linkage software package such as the generalized record linkage system maintained by Statistics Canada or a version currently in use at the Ontario Cancer Registry. The latter software is self contained and really runs in a PC environment. It can be purchased for a relatively modest amount of money which will be far less than the cost of developing a system from scratch in Ukraine. To date, there appears to be no experience or expertise in such record linkage in Ukraine. For example, the "linkages" conducted by Dr. Cortushin's staff for the thyroid project simply involved multiple passes of two files against each other and picking off "matches" on the basis of satisfying some arbitrary set of criteria. One pass is made for each set of criteria, all these potential matches are then given back to the endocrinology institute and how they are resolved is left to the discretion of the data processing people in the endocrinology institute who, again, have no experience in this area.

The small group training session could be held at Dr. Cortushin's institute in the fall, with participation by data processing personnel from the Institute of Radiation Medicine, the endocrinology institute, Dr. Cortushin's staff and data processing individuals from the Institute of Oncology, and any other organization which might be relevant. Thus, this would provide the ability to conduct such linkages in several institutions which may well be necessary for both the leukemia and thyroid projects. For example, it may be necessary to transfer the software expertise to conduct linkages against passport files, etc., as discussed above. This proposal would be developed more fully in the near future and submitted to NCI for their consideration.

D. Extraction of Information on Retrospective Cases of Leukemia and Lymphomas:

① It is planned during the pilot study to carry out the extraction of a sample of such cases from Dnepropetrovsk oblast. Apparently, the computerized cancer registration system has not yet started in this oblast, though it has started in several of the other oblasts which will be included in the full study, with data being extracted starting in various years. However, as indicated above, it is unclear that data from the Cancer Registry will be available for linkage purposes for the leukemia study.

② Dr. Gudzenko plans to visit the oncological dispensaries and hematologic departments of the remaining five oblasts during the pilot study to assess the state of records on historical leukemia/lymphoma cases. This probably should be done in conjunction with the obtaining of a sample of such cases for the pathology review (see below).

③ Dr. Gudzenko estimates that approximately 100 records of leukemia/lymphoma cases will be notified in Dnepropetrovsk oblast each year since 1987.

E. State Chernobyl Registry:

During the visit to Dr. Cortushin at the Chernobyl State Registry, the following points emerged:

① The computer equipment from NCI is unlikely to be installed and functional for at least 4-6 months. It seems apparent that no work on the leukemia project will be done by Dr. Cortushin and his staff until this equipment is installed.

② There seems to be no further progress in identifying and clarifying the various sources of data input into the register, other than that which comes directly from the individual oblast registries. It may, therefore, be unrealistic to expect that the representative nature of the registry in terms of all liquidators will ever be known with certainty. However, unless the registry was biased with respect both to dose and loss to follow-up, this should not affect the internal validity of the leukemia study which, of course, is the primary concern.

③ Apparently, files from both the military and the Ministry of Internal Affairs are currently physically at the Institute, but have not yet been integrated into the registry. However, it seems that no identifiers (e.g., names) are included in the military or Ministry of Internal Affairs files and, hence, such individuals cannot be included meaningfully in the study due to the inability to follow them directly by the means to be employed for those currently in the Chernobyl State Registry. It is also unclear when resources will be available to merge the three files together. My distinct impression from Dr. Cortushin was that this was on the "back burner," and it is not clear when or, frankly if, this will be achieved.

However, according to Dr. Vadim Chumak, there were three groups of military who participated in the cleanup. Only the "regular officers" are in the military file and not in the Chernobyl State Registry, with the two largest groups, namely, ordinary soldiers and reserve officers being contained in the State Registry. Dr. Chumak also thinks that these two latter groups are those primarily with any substantial doses. Hence, the loss of the military file may not lead to any serious loss of information. These facts should be clarified, and the actual numbers involved in each of the three groups identified, if possible.

④ We were told that passport numbers have not been included in the Chernobyl State Registry since 1992. It is not clear what this means, i.e., have all passport numbers been removed from the file or is it simply that no passport numbers have been collected from individuals registered since 1992? The latter would be relatively unimportant, but the former could be important.

⑤ As discussed above, it is clear that none of Dr. Cortushin's staff have any meaningful experience of record linkage. In particular, the concepts of probabilistic linkage, which I believe to be essential for the success of the leukemia study, are clearly foreign to them.

⑥ The Registry is currently in Foxpro Database format. Irena Gubina will be in charge of data processing at the Institute of Radiation Medicine's epidemiology group. She plans to maintain the

study cohort file in Foxpro, although she is also considering the possibility of using the ACCESS database system.

⑦ I promised to send Dr. Cortushin and his group material on probabilistic record linkage including the availability of software packages to conduct such linkages.

F. Leukemia Study Database Format:

① Discussions were held with Irena Gubina and Dr. Gudzenko about the possible database structure for the study. I suggested that it would be sensible to maintain two files, one on the full cohort and the other of the sub-cohort/cases. The latter file will, of course, contain substantially more information than the former. The database system to be used (Foxpro or ACCESS) should be chosen on the basis of convenience of usage and familiarity with those using the system in a hands-on fashion. However, one requirement which should be available in both databases is the ability to dump part or all of any records in a fixed length "flat file" in ASCII, since this will be the simplest way of interacting with other software packages, e.g., record linkage packages or statistical analysis software. This should not be a problem but will be very necessary since it will provide by far the easiest exchange of data between the database and such packages. For example, both the GRLS record linkage package and EPICURE which currently is the most appropriate package for fitting regression models to Poisson or Cox proportional hazards data require such flat files for data input. Further, for these purposes, records need only to be sequential and do not need to be accessed randomly.

② Dr. Gudzenko, in the second quarterly report, has laid out the information intended to go in the database. Although we did not review this in detail, I had two specific suggestions/questions. First, the passport numbers do not appear to be in the record structure, and Dr. Gudzenko agreed to add this to the record structure. Secondly, it is not clear whether items such as residence which, of course, will change from time to time for some cohort members, can be recorded in multiple records in the file, i.e., is a separate record of each residence or, alternatively, is only the current residence

kept in the file? This is very important since one of the criteria for eligibility should be residence at the time of first employment at Chernobyl, and it still is not clear to me that this is actually available on the State Registry file.

G. Protocol for Pathology Review During Pilot Study:

I reviewed the draft protocol for the pathology review prepared by Professor Burch and Dr. Gudzenko prior to my arrival. In general, this appeared to be appropriate, although I did make a few suggestions for changes, as follows:

① The random sampling procedure needs to be more clearly delineated. In particular, a random sample stratified by oblast and year of diagnosis in addition to diagnosis needs to be obtained. However, it will not be possible to stratify simultaneously on all three variables given the relatively small number of individuals to be selected and possibly stratification by year may have to be dropped. In this case the randomization process should be based on a list for each oblast in which each entry on the list has a random year of diagnosis and a random page number. This will be used to access the apparently sequential logs of cases kept by each oblast, oncological dispensary or hematologic department. Thus, for example, to obtain a case of AML, one takes the first item on the list, goes to the log for that year and the page number for that year, then starts half way down the page and moves forward sequentially until identifying the first case with the final diagnosis of AML which then satisfies the other requirements, i.e., male and of an appropriate birth year.

② Cases should be restricted to those aged between 20 and 40, between 1986 and 1990 so that age at diagnosis distribution should be comparable to that for the liquidators.

③ The actual process for the pathology review needs to be defined much more specifically in a protocol which should be developed in collaboration with the hematologists and pathologists involved. However, this process should include some measure of inter-observer variability, and possibly intra-observer variability (the latter being done by randomly including material from a

single case twice in the review with the reviewer being ignorant of and blinded to this fact. Professor Burch conducted a similar exercise, i.e., a pathology review of brain tumors from a number of different countries at the IARC in Lyon last year and his experience in this case should be invaluable.

④ I pointed out that the primary function of the review was a) to determine the availability of material and b) to give the review team some experience working together. The process will not, of course, be able to identify false negatives, i.e., any diagnoses missed by the original oblast procedure, except to the extent that, for example, leukemia cases received a related diagnosis which had been extracted for the purposes of the pathology review. False positives can, of course, be identified during the review process. Comparison of the review diagnosis with the original oblast diagnosis will only be important if there are, indeed, cases for which the bulk of diagnostic material is unavailable, in which case we will have to rely on the original diagnosis to confirm a retrospective case. However, the general opinion seems to be that such material should be available even for cases diagnosed as long ago as 1987.

⑤ I suggested that the time of the pathology review should be set reasonably far in the future, since in my experience collecting clinical pathological material always takes longer than expected. Further, the clinical records will need to be translated at a minimum into English for the U.S. pathologists and, possibly, also into French for the French pathologists.

⑥ Since Dr. Matsushima, the original Columbia-based leukemia pathologist has now resigned, I suggest that I consult with Dr. Finch to identify a suitable individual to fill the second slot on the U.S. side for the pathology review, preferably with an individual from the Columbia system; this will also be discussed with Dr. Shelanski, Chairman of the Department of Pathology at Columbia.

H. Dosimetry:

Discussions were held with Drs. Vadim Chumak and Gudzenko concerning the current status of dosimetry. The following points emerged:

① Dr. Chumak states that there are five approaches to estimating dose:

- Analytic Dose Reconstruction (ADR): This is a detailed reconstruction based on knowledge of radiation fields and exact time and location information from the individual liquidator. This is ***not*** a process which can be carried out based on the current IARC-designed questionnaire and, in Dr. Chumak's opinion, can only sensibly be carried out for staff of the Chernobyl nuclear power plant since these would be the only liquidators familiar enough with the local geography to give appropriate information.
- Fuzzy set doses (FSD): This is a very approximate method which gives an average dose to individuals based primarily on their occupation and time spent in the 30km zone. Again, in Dr. Chumak's opinion, this is ***all*** that can be done from the current IARC questionnaire, and he also emphasized that the methodology itself required substantially further work for which currently no funding is available.
- Electron Paramagnetic Resonance (EPR): This, of course, is only available for those with tooth samples and Dr. Chumak emphasized that, again, funds are not available to maintain the system for tooth collection amongst liquidators which he has initiated. He also clearly regards EPR as the "gold standard" as far as such exists.
- FISH: This appears to be the only dosimetric method that could be available for the entire subcohort and all prospective cases to the extent that resources and response rates permit. However, Dr. Chumak stated that for cases, even pre-treatment bloods may be inadequate to obtain decent samples for FISH.
- Official Doses: These, of course, are the current doses contained in the State Registry, but in Dr. Chumak's opinion, these will only be available for a minority of such subjects.

② Dr. Chumak suggested that there should be an immediate investigation as to the possibility of obtaining post-mortem teeth from all cases.

③ There seems to me a major potential problem in dosimetry in that there appears to be unlikely any consistent method of dose estimation for all the relevant study members, with the possible exception of FISH, which, according to Dr. Chumak, may have problems for dose estimation in cases even if pre-treatment bloods are obtained. An issue which I think needs addressing immediately is the availability of methods for utilizing data collected from a number of sources with varying measurement error structures and which could contain both systematic and random errors. For example, if a high proportion of cases have EPR but only a small fraction of the subcohort have EPR can methods be developed to obtain unbiased risk estimates? I suggest that I raise this question with Dr. Heitjan, and possibly some of his colleagues from the Biostatistics Department. Intuitively, it would seem that if one knew the error structure for the various approaches, one could overcome bias by appropriate analytic techniques, but this remains to be determined.

④ I told Dr. Chumak that, in my opinion, it was essential that all individual dose estimates by all procedures be included in the epidemiology database since one would need this information to conduct appropriate risk analyses and uncertainty analyses. Apparently, he had wanted initially only to give a composite "best estimate" of dose for each individual, but I believe I persuaded him that all estimates were necessary.

I. Overall Conclusions:

I believe that substantial progress has been made during the pilot study, and I find this encouraging.

I plan to extract from my above comments certain specific questions to which the answers are currently not available and give these in summary form to Dr. Gudzenko so that she can either ask these questions or obtain the relevant information. The top priority items, in my opinion, are to initiate the fieldwork in Dnepropetrovsk oblast, which should be done as soon as possible, and to

pursue the question of the appropriate method of using different dose estimates from different sources for different members of the subcohort and cases in order to obtain unbiased risk estimates. As stated, I plan to initiate this process on my return to Columbia as soon as possible.

III. Ukranian Thyroid Study:

This was my first visit for a number of years to the Institute of Endocrinology in Kiev in connection with the study of thyroid cancer. Overall, my impression was that a good start had been made to the study and, although a number of issues were identified during my visit which require attention, the study appears to have been initiated satisfactorily and has the potential for yielding meaningful epidemiologic data on the relationship between thyroid disease and exposure, primarily to 131_I.

The areas on which I focused and on which I provide my detailed comments below, were: a) the definition, location and recruitment of the cohort.; b) issues relating to the various questionnaires; c) coding and data entry; and d) the possibility of a preliminary statistical analysis based on a small number of thyroid cancer cases. My detailed comments follow.

A. Definition, Location and Recruitment of the Cohort:

① In terms of defining the basic cohort to be included in the study, there are two fundamental requirements. First, the selection process should be such so as not to introduce any meaningful biases in the risk estimates and, secondly, the cohort should be of adequate size, i.e., with an expectation of a sufficient number of cases to yield adequate power to provide risk estimates with sufficiently narrow confidence intervals to address questions such as the relative biological efficiency of ^{131}I and external gamma radiation in inducing thyroid cancer.

With respect to bias, certain principles need to be borne in mind. These principles are, of course, well established in the epidemiologic literature, but do have immediate and important practical consequences for choosing the cohort. The most immediate practical consequence is that if, of the initially selected 20,000 cohort, only 50% can be traced and included in the study, this does not

necessarily affect the results of the study, and one may validly go on and locate a further 50% of the next 20,000 and so on until the required sample size is reached to provide the necessary power. Unless the sampling is biased with respect to the combination of dose and disease risk, failure to locate a substantial proportion of the initially selected cohort would have no impact upon the relative risk estimate although potentially a greater influence on excess risk estimates. The reality may be such that, indeed, one has to live with substantially lower location and response rates than would be ideal.

Virtually all cohort studies are never truly representative of the target population because, inevitably, there will be a substantial non-response rate; this is true, for example, of such well known cohort studies as the British doctors' study (response rate about 50%) and the nurses' cohort study (response rate about 20%). Yet the results of these two studies are universally accepted as valid. The issue of representativeness relates only to the so-called external validity of the study, i.e., can the results be applied to all individuals; in general, there is no reason to expect that results from internally valid studies will not be applicable to the general population, particularly when results are presented specifically for factors such as age at exposure and gender.

In summary, I would suggest that the most important thing is to build an adequate sized cohort and not worry too much about its representativeness; the efforts should be put into obtaining enough people and excessive effort should not be put into exhaustively trying to trace all members of the initially identified cohort.

② The present study presents an ideal opportunity to conduct either case-control or case-cohort within cohort studies. This is not analogous to the leukemia study where, in fact, the use of such designs was based upon cost considerations, where these study designs have the same potential biases as conventional case-control studies. In the present situation, i.e., the thyroid study, the case-control and case-cohort within a cohort study design would not be subject to the biases of conventional case-control studies, but would provide the same degree of bias as the full cohort study

but at a fraction of the cost. In particular, such a design would involve carrying out laboratory processing of blood samples and working up individual doses ***only*** for cases and a sample of the full cohort. For example, if laboratory workup and dosimetric workup were carried out for 5,000 instead of 20,000 individuals, costs for these activities would be reduced by about 75%, without any meaningful loss of efficiency. Questionnaires and blood samples would, of course, have to be stored so that they could be retrieved for cases as they occur, but I assume this will be done anyway. I realize that it might be hard to sell this concept politically, both in the U.S. and Ukraine, but it would lead to very substantial savings on the resources required; alternatively, the same resources could be spent but the size of the cohort could be substantially enlarged and, hence, the number of cases could be substantially increased and, hence, the efficiency could be substantially improved.

③ Currently two sources are being used to locate the cohort. "Record linkage" has been carried out between the dose file of Dr. Likhtarev and the Ukrainian State Chernobyl Registry. In fact, this is not a linkage in the true sense since it simply involves repeated passes of the two files against each other using a different series of criteria for exact matching without any consideration either of the relative frequency of the identifying items, nor rates of errors in recording data. The process by which the final "matches" were resolved at the Institute of Endocrinology is not documented, nor could it be described by the individuals from the DCC.

The results of this "linkage" are, to put it mildly, dubious. The table attached shows the percentage of cases on the dose file "matched" to individuals in the State Registry. Even though the great majority of the Likhtarev subjects are supposed to be on the Registry, the overall matching rate, i.e., 24% is extremely low. Further, it varies substantially from raion to raion varying from 4% to 51%. The problem could be in the matching process used, or could be in the data on the Registry, or a combination of the two.

I suggest that the first step would be to introduce, as soon as possible, probabilistic record linkage into the state registry and re-run the linkages in order to minimize any problems arising from the

be utilized to increase this rate, or should a further sample of the full potential cohort of 80,000 be selected and the two existing tracing methods continue to be used as the only methods. I think this would best be answered after the repetition of linkage using probabilistic linkage approaches at the State Registry, since, hopefully, this might improve the overall tracing rate. I think it worthwhile to pursue the possibility of using resources such as the passport file at this stage, but would hesitate to recommend their full-blown use until the linkage has been repeated.

In order to estimate the final size of the cohort, it is necessary, of course, to estimate not only the location rate of study subjects, but also their response to the invitation to screening. The only data which I have available at the moment are from the one mobile team whom we met with in Kiev, who stated that 53 of the 90 invited individuals had actually turned up, i.e., a participation rate of 59%. If this is multiplied by the approximate estimated location rate described above, we obtain an overall inclusion rate of 38% based on the approaches used so far. This would mean that approximately 11,600 of the 20,000 initially selected cohort would participate in the study. If we were to multiply this by four (given the size of the Likhtarev file of 80,000) we would, thus, recruit about 40,000 study subjects into the final cohort. It is probably worthwhile recalculating the power of the study based on this number or, preferably, based on the combination of this number with the data from Belarus, before decisions can be made as to using alternative methods of location or improved methods for study participation, e.g., giving gifts to those who participate.

④ During the visit to Kiev, various approaches to inviting study subjects to participate were considered. I would strongly recommend that only a single letter be sent to potential study subjects (i.e., the letter that is currently sent) coming from the Ministry of Public Health, outlining the study and that such a letter should be sent, ideally, several weeks prior to the visit by the local medical staff. I see absolutely no point in including the postcard for return since all this does is to identify the extremely small number of subjects who will send the postcard back saying that they do not wish to participate (4 out of 594 in the experience to date). Many people simply do not return the postcard, but still participate in the study. The letter would then be followed by the personal visit

of the local medical staff asking the individual to participate and scheduling a suitable date for the visit. I note, at the moment, the scheduling is done entirely by the local medical staff which is not ideal, but there seems to be little practical alternative, and so far it appears as though the system has worked, i.e., sufficient study subjects are recruited to keep the mobile teams busy, but not too many are recruited to flood the resources of the mobile teams.

⑤ The issue which has thus far appeared not to have received too much attention is the necessity for inviting people back for the second and subsequent screenings. I would suggest that information be obtained at the time of the first screening on names and addresses of subjects' parents, and other relatives or potential contact persons such as friends. It might also be useful to attempt, if possible, to get some information on any expected plans in two to three years time for the study subject, e.g., does he/she anticipate going away to some educational institute at this time, do they have such institutes in mind, etc. These procedures are fairly standard in cohort studies where you need repeated contact with study subjects, and obviously need to be initiated now, otherwise future contact is liable to be problematic.

B. Issues Relating to the Various Questionnaires:

① The current questionnaires need modifying in several ways. First, it will be important to identify the source of information, i.e., parents, child, or combination of the two. Although this can be done on a question-by-question basis, it seems more practical to allow the interviewer at the end of the interview to characterize the interview into, say, five categories, i.e., all responses provided by parents, most responses provided by parents, responses equally provided by parents and subject, most responses provided by subject, all responses provided by subject. This should be adequate for any statistical analysis. Secondly, there should be some semi-quantitative estimate of the "reliability" of the interview provided by the interviewer. Again, this could be a five-category scale, ranging from "seems very reliable" to "seems very unreliable." Although this inevitably is somewhat subjective, it has proved useful in other studies, particularly to eliminate what appear to be unsatisfactory interviews.

② There needs to be a specific policy as to the age at which information can be requested solely from the study subjects, e.g., from those who were age 10 or 12 or more at the time of the accident. Information on anyone under this age which is not obtained with the parents' cooperation probably should not be used.

③ Attention needs to be paid urgently to the issue of setting up appropriate coding boxes and codes on the questionnaire; it appears as though much of the coding could be done directly by interviewers, e.g., sex, etc. The boxes need to be set on the right hand side of the questionnaire for ease of data entry, and boxes which will be filled in by the interviewer as opposed to those by coding staff, should be distinguished from each other, e.g., by shading those boxes to be completed by coding staff.

④ Finally, there needs to be an ongoing assessment of the questionnaire, e.g., by routine meetings of interviewers to see whether certain specific problems are regularly emerging during interviews, so that if need be the questionnaire can be corrected.

C. Coding and Data Entry:

① This process needs to be initiated as soon as possible. Without coding and data entry quality assurance for the questionnaire data will be hard to tabulate and examine. In particular, a coding manual needs to be developed as soon as possible in collaboration with those who have already conducted interviews and who may well help to contribute appropriate suggestions for coding.

② Coding procedures need to be standardized, e.g., using the same codes for the same responses, e.g., 1 for "yes," 2 for "no," 9 for "don't know" (don't know should ***not*** be coded as a blank since one cannot distinguish this from simply a failure to code).

③ There should be two files of data for study subjects. The first is an "administrative file" on which ***all*** potential study participants are included. Currently this would include the 20,000 selected

from Likhtarev's file. The information to be contained in this file would include identifying data, study I.D., information from the case recruitment sheet, date of first screening, availability of information, of laboratory data, dosimetric data, detection of any thyroid pathology, etc. (the latter field could simply be scored yes/no). The file would also contain information on second and subsequent screenings, e.g., letter sent to invite for second screening and the date sent, date of second screening, etc. This administrative file will be the basis for tabulating response rates, keeping overall track of the study, generating letters of invitation, etc. The second file will be the study database containing the detailed information from the questionnaire, detailed laboratory results, detailed clinical data, etc. This of course would need to be set up in a standard database format since it will contain repeated segments of information from each screening visit. Arguably, the highest priority should be given to the establishment of the administrative file since it is likely to be easier, and the information will more urgently be needed.

④ Consideration needs to be given to the best method of interfacing the database for the study file to other software, e.g., SAS and other analytic software programs.

D. Preliminary Estimation of Risks:

It has been suggested that risk estimates could be derived from the study using only a small number of cases when available, e.g., 20-30. I recognize the desirability of obtaining initial estimates from the study as soon as possible in order to underpin its feasibility and utility. However, I am very concerned that such estimates and the manner in which it is currently proposed to obtain them could be misleading and substantially weaken the perceived credibility of the study.

My first concern is that any estimates will inevitably have very wide confidence intervals and could, for example, conceivably give rise to estimates of negative risk coefficients; if they are positive it is certainly possible that the confidence interval could include an rbe of 0.1or 10.0 compared to risks for external gamma radiation.

Secondly, the difficulty of fitting the data does not seem to be fully appreciated by Dr. Likhtarev's team who plan to carry out risk estimation. The model to be fitted would be the linear excess relative risk model fitted to proportional hazards data, and the only software that I am aware of currently available to fit such models is the PEANUTS module of the package EPICURE which produces likelihood based point estimates and interval estimates. However, the program, to the best of my knowledge, it is not currently available in Ukraine, and requires appreciation of the subtleties involved in fitting such models. Fitting other statistical models using approximate methods is likely to lead to very misleading estimates. A simple example is that if confidence intervals are based on the ratio of the point estimate to its standard error, which is often done, the resulting confidence intervals are completely erroneous.

The third, and perhaps the most worrying possibility, is that the selected cases and selected individuals with dose measurements could be a highly biased sample in some way which might not be recognized by the dosimetrists. This, of course, would result in meaningless risk estimates.

As I say, I understand the desirability of producing some risk estimates, and that Dr. Likhtarev's group feel that they should be the individuals to do it given their ownership of the dose data. However, I would argue most strongly that any exercise of this nature needs to be undertaken most cautiously and would need the involvement of an individual with substantial experience in conducting risk estimation in radiation cohort studies. I would be most happy to offer my own services to Dr. Likhtarev's group in this respect, and would expect no recognition of any contribution I might make so this would not detract in any way from credit going to him and his group. However, I realize it might be very difficult to get Dr. Likhtarev to accept this solution, but as I say, I think we have a major potential problem here.

State of the 20.000-cohort for 01.06.98

Raion		Found			%	
	Total	in Chern.Reg	Manual search	Total	in Chern.Reg	Manual search
Chornobyl	1484					
city Prypyat	1584					
Polisya	1399	511			36,53	
Ivankiv	787	378	297	494	50,61	40,30
Kiev oblast	**5204**	**884**	**297**		**16,99**	**5,71**
Kozelets	2089	215	1156		10,29	55,34
Repki	1377	572	793		41,54	57,59
Chernigov	2858	997	1386		34,88	48,50
city Chernigov	1179	51	1024		4,33	86,85
Chernihiv oblast	**7503**	**1835**	**4359**		**24,46**	**58,10**
Narodichi	4279	811	1256		18,95	29,35
Ovruch	3072	714	2157		23,24	70,21
Zhytomyr oblast	**7351**	**1525**	**3413**		**20,75**	**46,43**
city Kiev		**542**	**413**			
Total	**20058**	**4786**	**8482**		**23,86**	**42,29**